

В.В. Скалозуб, В.М. Горячкін, І.В. Клименко, І.А. Терлецький
**ДОСЛІДЖЕННЯ ІНТЕЛЕКТУАЛЬНИХ МОДЕЛЕЙ УПРАВЛІННЯ
НА ОСНОВІ ПРОЦЕДУР КЛАСИФІКАЦІЇ НЕВИЗНАЧЕНИХ ДАНИХ
ЗІ ВСТАНОВЛЕНИМИ ВИМОГАМИ ДОСТОВІРНОСТІ РЕЗУЛЬТАТІВ**

Анотація. Стаття присвячена дослідженням властивостей і розвитку інтелектуальних моделей управління складними системами за умов невизначеності даних на основі процедур класифікації на основі методів редукції та статистики каппа Коена. Застосування цих методів забезпечує достовірне вирішення завдань з урахуванням оцінки граничної розмірності моделей класифікації. В роботі були досліджені можливості удосконалення нейронних мереж Хеммінга для класифікації даних у форматах нечітких величин і certainty factor CF(A). Також були визначені особливості удосконаленої математичної моделі завдань нечіткої класифікації на основі набору шаблонів ознак. Представлено структуру програмного комплексу інформаційної технології управління призначенням/відбором виконавців на основі класифікації наборів шаблонів із певних нечітких ознак, який використовує процедури редукції і каппа статистики.

Ключові слова: класифікація наборів шаблонів, невизначені дані, редукція розмірності, статистика каппа Коена, нечіткі величини, certainty factor CF(A), модифікована мережа Хеммінга, інформаційна технологія, програмне забезпечення, завдання визначення авторства та призначення/відбір виконавця.

Вступ та постановка проблеми. Натепер засобами широкого кола технологій класифікації та діагностування неповно визначених даних щодо станів та умов функціонування складних систем вирішуються завдання вибору варіантів керування різноманітними технологічними процесами, відбору виконавців встановлених завдань тощо [1, 5, 6, 9]. При цьому актуальними залишаються завдання формування адекватних математичних моделей процедур класифікації та встановлення їх коректності, повноти та достовірності результатів. У статті [1] отримано розвиток моделей класифікації в умовах невизначеності даних з використанням процедур редукції та статистики «каппа Коена» [4, 8]. В ній розглянуті завдання достовірності результатів класифікації. А саме, забезпечення узгодженості ймовірнісних вимог щодо достовірності наступних результатів, розмірності параметрів моделей

(шаблонів), а також кількості даних (класів) за якими виконується діагностування. Для реалізації процедур класифікації запропоновані модифіковані мережі Хеммінга (МХН) [1, 2, 10], а для забезпечення встановлених вимог достовірності результатів – методи редукції, пристосовані для завдань класифікації за рахунок використання статистики каппа Коена. Коректність, достовірність та ефективність моделей і процедур класифікації при невизначених даних [1, 2] визначалась на основі результатів числового моделювання. Разом з тим потребують подальшого дослідження завдання щодо уточнення і розширення метрик нечіткої подібності шаблонів класифікації, а також застосування метрик формату $CF(A) \in [-1; 1]$ (біполярні коефіцієнти, certainty factor), формування інтелектуальних засобів для управління на основі класифікації невизначених даних складних систем і процесів тощо. Таким завданням присвячена представлена стаття.

Аналіз останніх досліджень і публікацій. У статтях [1, 2, 5, 9] і багатьох інших відзначаються можливості значної ступені існування джерел і умов невизначеності щодо характеристик та станів сучасних складних систем, наявності різноманітності метрик невизначеності, повноти отримання даних тощо. Серед інформаційної технології (ІТ), які використовують моделі класифікації та діагностування, відзначимо процеси оптимізації потоків замовлень у сервісних системах (С&С) [2, 6], розформування-формування вантажних залізничних составів та керування певними процесами [3, 9].

Розвитку моделей та багатопараметричних інтелектуальних процедур класифікації та діагностування за неповними і збуреними даними присвячена стаття [1]. В ній подано нові постановки, математичні моделі та реалізації завдань класифікації за недетермінованими даними. Одним із прикладів моделі класифікації за нечіткими даними являється окреме завдання, в якому на основі нечіткої класифікації встановлюються автори україномовних текстів. Також на основі моделей класифікації даних у форматі коефіцієнтів упевненості $CF(A)$ [5, 6] виконується оптимальний відбір кандидата на певну виробничу діяльність (ВД). При тому модель ознак кандидатів на ВД має удосконалену багатовекторну структуру. У всіх зазначених процедурах передбачене застосування методу редукції простору моделі [3] за рахунок дослідження статистики каппа Коена [4, 8], що гарантує отримання результатів із встановленою достовірністю результатів вибору.

Завдання нечіткої класифікації (також і класифікації за ознаками метрики $CF(A)$) реалізується шляхом застосування модифікованої моделі асоціативної

пам'яті Хеммінга (МХН). В МХН величини відстані між нечіткими елементами (μX) зразків моделі класифікації (w_{ik}) і вхідним вектором $X = \{x_i: i=0\dots n-1\}$ визначаються на основі нечіткого відношення [9] виду

$$R(W, X) = \{\mu R(w_i, x_i) / \{w_i, x_i\} = (1 - \text{abs}(w_i - x_i)) / \{w_i, x_i\}\}, i=0, 1, \dots, n-1. \quad (1)$$

Відношення (1) дозволяє вимірювати відстані між нечіткими векторами $X = \{x_i: i=0\dots n-1\}$ та відповідно моделлю шаблонів класифікації (w_{ik}). Зауважимо, що при однакових значеннях ступенів приналежності величин $\{w_i, x_i\}$ значення $\mu R(w_i, x_i) = 1$. Якщо одна з величин $\{w_i, x_i\}$ дорівнює 0, а інша 1, тоді значення $\mu R(w_i, x_i) = 0$, інакше $\mu R(w_i, x_i) \rightarrow [0; 1]$. Відношення (1) дозволяє модифікувати МХН для нечітких моделей класифікації, а також моделей метрики $CF(A)$ [5].

Проблематика щодо визначення розмірності параметрів математичних моделей із застосуванням процедур індукції і редукції представлена у [3]. В якості процедури, за якою виконується формування простору математичних моделей, призначених для завдань представлення «закономірностей рівності» ($y=f(X)$), використовують α -процедуру. Відзначимо, що завдання класифікації відноситься до визначення «закономірностей схожості», для реалізації яких і була запропонована спеціалізована процедура на основі використання статистики каппа Коена [1]. При тому оцінки граничної розмірності параметрів моделі класифікації, що синтезується « n_0 », визначаються відповідно до рівнянь

$$n_0 = (\varepsilon * L + \ln(h)) / \ln(m). \quad (2)$$

Відповідність даних рівнянням (2) визначає умову щодо виконання заданих вимог достовірності « ε, h »; $(1 - h)$ оцінка ймовірності безпомилкового розділення випадкової і незалежної вибірки довжини « L » при заданій граничній величині помилкової класифікації « ε » [3].

Призначення статистики «каппа Коена» [4, 8] (у пропонованій процедурі редукції розмірності математичної моделі класифікації) полягає у встановленні «подібності», або належності конкуруючих багатопараметричних математичних моделей до одного класу, щодо відповідності вимогам шаблонів, сформованих із ознак властивостей контрольованих процесів. Для такої перевірки «подібності» необхідно утворити певну множину математичних моделей, шаблонів із різними наборами змінних, для яких в подальшому виконувати порівняльний аналіз їх відповідності між собою. Наприклад, в [1] за оцінками каппа статистики перевірялась «подібність» результатів класифікації для всіх пар змінних шаблонів (X/X_j) та (X/X_i). При цьому за каппа-оцінками наборам змінних (X/X_j) для кожного шаблону приписувалися значення «+» або

«-» в залежності від результатів класифікації. У табл. 1 приведено структуру моделі класифікації та її умовні результати для шаблонів з шести змінних. Далі за порівняльним аналізом результатів класифікації наборів змінних формуються таблиці розбіжностей та розраховують статистичні оцінки «каппа Коена» [4]:

$$K = (P_0 - P_e) / (1 - P_e), \quad (3)$$

де P_0 – імовірнісна оцінка що показує наскільки спостережувана узгодженість краща за випадкову, а P_e – результат підрахунку максимально можливої узгодженості за винятком випадкової узгодженості конкуруючих шаблонів.

Таблиця 1

Класифікації шаблонів з параметрів (X1, X2, ..., X6).

K_i	L_i	X	X/{X ₁ }	X/{X ₂ }	X/{X ₃ }	X/{X ₄ }	X/{X ₅ }	X/{X ₆ }
K_1	L_{11}		+					+
	L_{12}	+		+		+		
	L_{13}					+		+
K_2	L_{21}					+		+
	L_{22}		+	+			+	
	L_{23}				+			
	L_{24}	+				+		
K_3	L_{31}		+	+				
	L_{32}	+					+	
K_4	L_{41}							
	L_{42}		+		+	+		
	L_{43}							
	L_{44}	+		+		+		+
	L_{45}					+		
K_P	L_{P1}	+				+		
	L_{P2}	+	+	+				

У таблиці позначено: K_i – класи/зразки моделі (можливі кілька екземплярів L_i); X – параметри моделі на поточному етапі редукції; X/{X_i} – скорочені набори моделей класифікації (без множини параметрів {X_i}); знаки «+» – визначення шаблонів переможців (класів) за моделлю МХН. У стовпці X позначаються шаблони-переможці для X наборів параметрів. Множина варіантів моделей класифікації (KM) містить всі комбінації пар скорочених мо-

делей $X/\{X_j\}$. Завдання полягало у визначенні серед скорочених моделей шаблонів $X/\{X_j\}$ такого набору параметрів $\{X_j\}$, які необхідно видалити на наступному етапі процедури редукції. Разом з тим можливості формування множин $\{X_j\}$ у процедурах редукції залишилися не розкритими.

Наведені вище результати свідчать про те, що формування коректних і ефективних інтелектуальних засобів технологій і систем управління на основі класифікації невизначених даних являється актуальним. Також актуальними для цього є дослідження можливостей метрик нечіткої класифікації та метрик формату $CF(A)$, як і розробка автоматизованих засобів реалізації процедур редукції.

Мета дослідження – розвиток інтелектуальних моделей управління складними системами за умов невизначеності даних на основі процедур класифікації, а також шляхом формування програмних засобів, які забезпечують достовірне вирішення завдань класифікації з урахуванням рівнянь (2) за даними метрик нечітких величин і коефіцієнтів впевненості $CF(A)$.

Результати та основний матеріал дослідження. Серед завдань цієї статті визначено дослідження можливостей метрик нечіткої класифікації та формату certainty factor $CF(A)$ щодо їх застосування для модифікування мережі Хеммінга (МХ) із областю значень векторів ознак $\{-1; +1\}$ для їх переходу до мереж МХН із областями нечітких значень μ_x ($X \rightarrow [0; 1]$), а також показників $CF(A)$ із значеннями множини $[-1; +1]$. В якості обґрунтування запропонованого нами рішення щодо МХН (1) розглянемо основні типи нечітких метрик Хеммінга відповідно [2], а також їх співвідношення з мережею МХ з ознаками $\{-1; +1\}$. В статті [2] метрика нечіткої відстані Хеммінга (з додатним відхиленням при введенні α_r) між нечіткими множинами $A \sim$ і $B \sim$, визначається за формулою:

$$d^+(A \sim, B \sim) = \sum_{r=1}^q \alpha_r \cdot |\mu_{A \sim}(u_r) - \mu_{B \sim}(u_r)| \quad (4)$$

де $u_r \in U$, $\mu_{A \sim}(u_r), \mu_{B \sim}(u_r) \in [0, 1]$, $r = \underline{1}, q$, коли α_r – індикатор:

$$\alpha_r = \{1, \text{якщо } \mu_{A \sim}(u_r) \geq \mu_{B \sim}(u_r) \text{ } 0, \text{якщо } \mu_{A \sim}(u_r) < \mu_{B \sim}(u_r) \} \quad (5)$$

Зауважимо, що метрика (4) з індикатором (5) дозволяє також розрізняти вимоги щодо переваги показників $\mu_{A \sim}(u_r), \mu_{B \sim}(u_r) \in [0, 1]$, тобто, вводити певні вимоги/умови стосовно оцінок порівнюваних нечітких векторів $A \sim$ і $B \sim$.

Метрика (4) не відповідає векторам ознак мережі Хеммінга МХ з областю значень $\{-1; +1\}$, тому в [6] була запропонована процедура перекодування показників $\mu_X: X \rightarrow [0; 1]$ до множини значень моделі МХ. При цьому відзначається принципова потенціальна можливість представлення однаковими кодами певних різних нечітких величин. Також відзначається необхідність поєднувати кодування з метою застосування моделі МХН, з процедурою класифікації зразків. Числові розрахунки на основі МХ показали існування таких же питань щодо реалізації завдань класифікації даних формату certainty factor CF(A) на основі використання подібних (4) метрик.

Дослідження процедур МХН на основі (1) підтвердили коректність і ефективність його застосування для модифікування МХН, як для нечітких метрик моделей класифікації, так і для моделей метрики CF(A). При тому також встановлені можливості визначення додаткових вимоги/умови щодо показників близькості порівнюваних нечітких векторів $A \sim i B \sim$. Простий змістовний приклад таких додаткових умов полягає в наступному. Нехай існують дві пари векторів нечітких оцінок певних ознак системи, які мають наступні значення $(\mu_{A \sim}(u_r) = 0.2, \mu_{A \sim}(u_k) = 0.8)$ та $(\mu_{B \sim}(u_r) = 0.3, \mu_{B \sim}(u_k) = 0.9)$. Тут абсолютні відстані між u_r та u_k дорівнюють 0.1, але відносні $-(\mu_{A \sim}(u_r) - \mu_{B \sim}(u_r)) / (\mu_{A \sim}(u_r); \mu_{B \sim}(u_r))$ суттєво різні для u_r та u_k . Для забезпечення можливості встановлення нових вимог до метрики відстані відношення виду (1) треба замінити наступним

$$R(W, X) = \{\mu R(w_i, x_i) / \{w_i, x_i\} = (1 - \text{abs}(w_i - x_i)) / \max\{\text{abs}(w_i); \text{abs}(x_i)\}\}, i=0, 1, \dots, n-1 \quad (6)$$

Як і (1), відношення (6) придатне для нечітких моделей класифікації, а також моделей типу certainty factor CF(A).

Приведемо деякі результати застосування модифікованих моделей класифікації МХН для зазначених типів даних. На рис. 1 подані приклади завдань

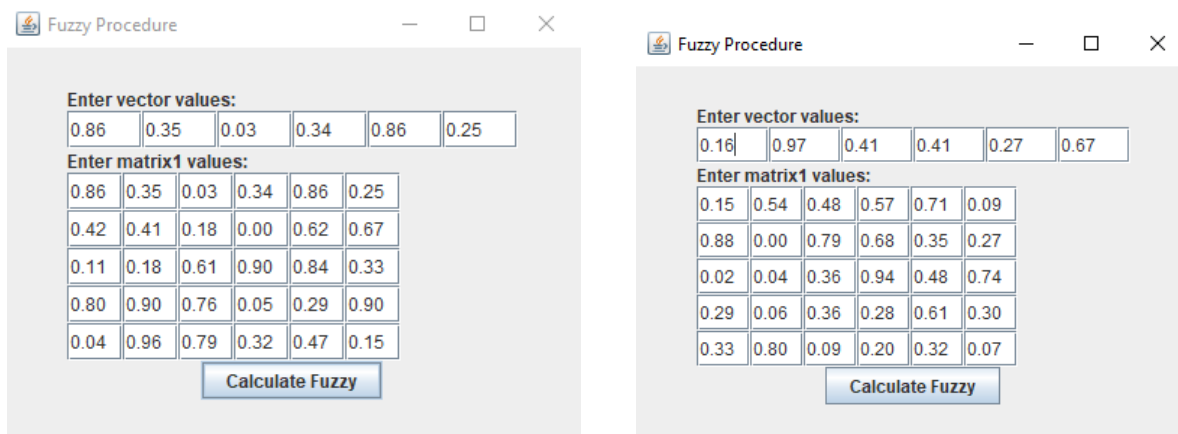


Рисунок 1 - Постановки завдань нечіткої класифікації

нечіткої класифікації, а на рис. 2 їх реалізація. В них, рис. 1 ліворуч, вхідний вектор відповідає шаблону (перший рядок матриці, індекс 1), а праворуч приведена реалізація моделі досить невизначеної нечіткої структури. Перше завдання було реалізоване за одну ітерацію, як у стандартній МХ, а друге – через кілька кроків процедури (індекс 5). При тому показник «value», що означає оцінку ступеня порівняльної достовірності результату класифікації, у другому завданні був досить близький до нуля (границі припинення відбору шаблонів МХ). На рис. 3 приведено приклад реалізації завдання нечіткої класифікації за МХН, з високим ступенем достовірності (показник «value»).

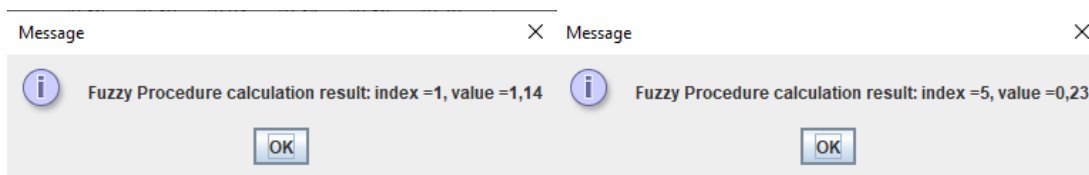


Рисунок 2 - Результати реалізації завдань нечіткої класифікації рис. 1

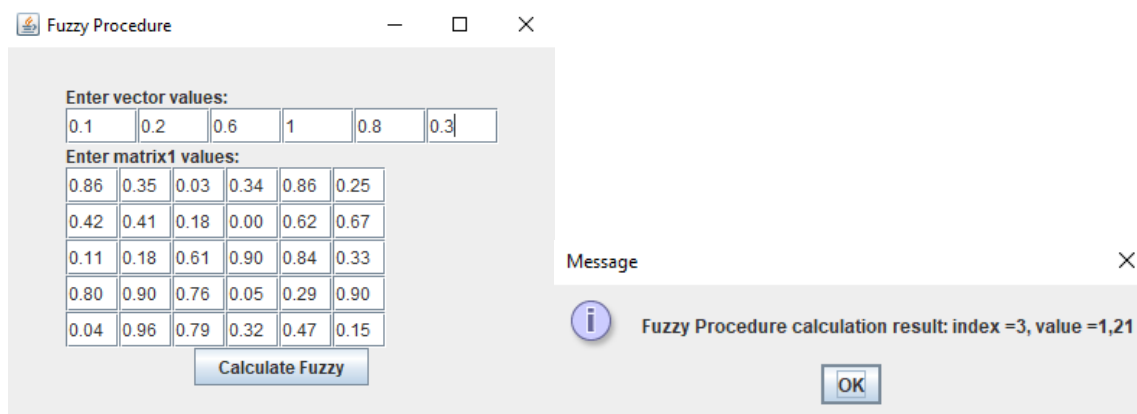


Рисунок 3 - Результати реалізації нечіткої класифікації за МХН для шаблону «3»

На рис. 4 та рис. 5 подано подібні результати класифікації за даними моделей типу certainty factor CF(A). Рис. 4 демонструє результат, коли вхідний вектор відповідає шаблону «5», а на рис. 5 досліджується довільний вектор ознак.

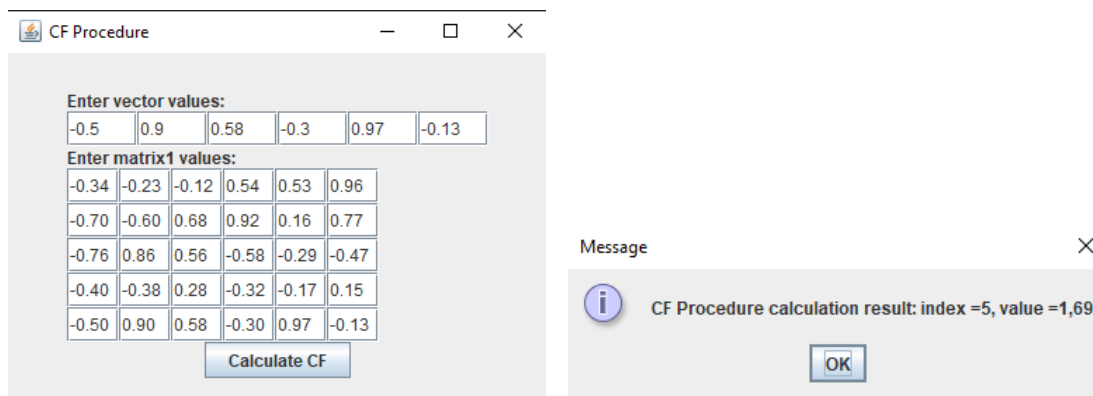


Рисунок 4 - Результати реалізації CF(A) класифікації за МНХ шаблону «5»

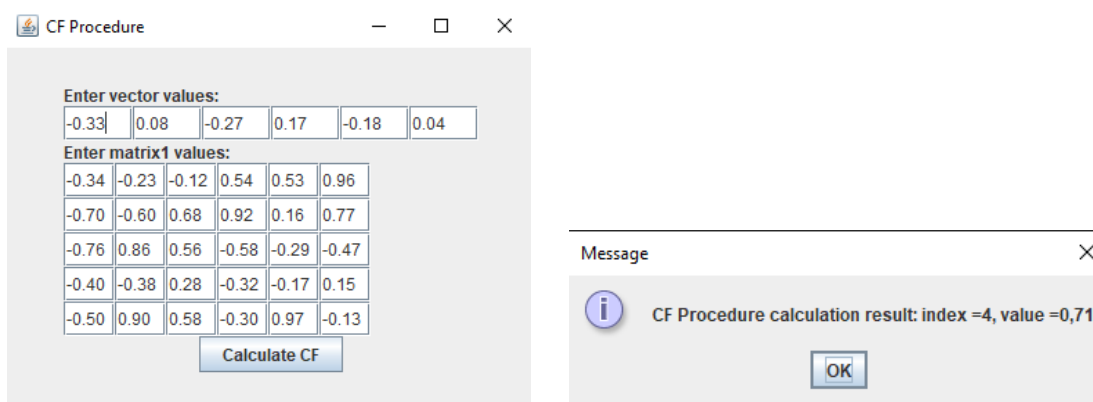


Рисунок 5 -Результати реалізації моделей завдань CF(A) класифікації

Приведемо сутність та результати виконання процедури редукції на основі каппа статистики ПКР (3), призначеної для формування достовірних математичних моделей класифікації за МХН, яка має такі етапи:

- 1) на основі моделей шаблонів з різним числом або складом параметрів визначити оцінку показника «каппа» (3),
- 2) якщо результати класифікації для різних моделей-шаблонів відповідають вимогам «подібності», скоротити модель, залишити один із шаблонів,
- 3) враховуючи значення « n_0 » (2), визначити і видалити найменше значимі чи найбільше «подібні» між собою параметри моделі класифікації.

Для пояснення сутності процедури редукції, запропонованій у [1, 6], використаємо умовний приклад табл. 2 зі структурою моделі класифікації

шаблонів з параметрами (X_1, X_2, \dots, X_6) табл. 1. У табл.2 приведені значення (3), отримані при співставленні узгодженості приведених у таблиці пар змінних.

Таблиця 2

Розрахунки при $m=5$ за зразками з урахуванням відомих даних X (KM_2, C_1)

$K(X_2/X_3) \quad 0,150$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>12</td> <td>5</td> </tr> <tr> <th>ні</th> <td>5</td> <td>4</td> </tr> </tbody> </table>		Так	ні	так	12	5	ні	5	4	$K(X_2/X_4) \quad 0,103$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>13</td> <td>4</td> </tr> <tr> <th>ні</th> <td>6</td> <td>3</td> </tr> </tbody> </table>		Так	ні	так	13	4	ні	6	3
	Так	ні																			
так	12	5																			
ні	5	4																			
	Так	ні																			
так	13	4																			
ні	6	3																			
$K(X_2/X_5) \quad 0,150$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>12</td> <td>5</td> </tr> <tr> <th>ні</th> <td>5</td> <td>4</td> </tr> </tbody> </table>		Так	ні	так	12	5	ні	5	4	$K(X_2/X_6) \quad 0,023$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>8</td> <td>9</td> </tr> <tr> <th>ні</th> <td>4</td> <td>5</td> </tr> </tbody> </table>		Так	ні	так	8	9	ні	4	5
	Так	ні																			
так	12	5																			
ні	5	4																			
	Так	ні																			
так	8	9																			
ні	4	5																			
$K(X_3/X_4) \quad 0,103$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>13</td> <td>4</td> </tr> <tr> <th>ні</th> <td>6</td> <td>3</td> </tr> </tbody> </table>		Так	ні	так	13	4	ні	6	3	$K(X_3/X_5) \quad 0,320$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>13</td> <td>4</td> </tr> <tr> <th>ні</th> <td>4</td> <td>5</td> </tr> </tbody> </table>		Так	ні	так	13	4	ні	4	5
	Так	ні																			
так	13	4																			
ні	6	3																			
	Так	ні																			
так	13	4																			
ні	4	5																			
$K(X_3/X_6) \quad 0,324$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>10</td> <td>7</td> </tr> <tr> <th>ні</th> <td>2</td> <td>7</td> </tr> </tbody> </table>		Так	ні	так	10	7	ні	2	7	$K(X_4/X_5) \quad 0,283$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>14</td> <td>5</td> </tr> <tr> <th>ні</th> <td>3</td> <td>4</td> </tr> </tbody> </table>		Так	ні	так	14	5	ні	3	4
	Так	ні																			
так	10	7																			
ні	2	7																			
	Так	ні																			
так	14	5																			
ні	3	4																			
$K(X_4/X_6) \quad 0,331$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>11</td> <td>8</td> </tr> <tr> <th>ні</th> <td>1</td> <td>6</td> </tr> </tbody> </table>		Так	ні	так	11	8	ні	1	6	$K(X_5/X_6) \quad 0,173$	<table border="1"> <thead> <tr> <th></th> <th>Так</th> <th>ні</th> </tr> </thead> <tbody> <tr> <th>так</th> <td>9</td> <td>8</td> </tr> <tr> <th>ні</th> <td>3</td> <td>6</td> </tr> </tbody> </table>		Так	ні	так	9	8	ні	3	6
	Так	ні																			
так	11	8																			
ні	1	6																			
	Так	ні																			
так	9	8																			
ні	3	6																			

Найвища в табл. 2 оцінка коефіцієнту «каппа» (3) у пари параметрів $K(X_4/X_6) = 0,331$ (помірний рівень узгодженості), а у пар $K(X_3/X_6) = 0,324$ і $K(X_5/X_6) = 0,173$. При цьому видаляється параметр X_6 через те, що «вплив» фактору X_6 на модель класифікації враховують та опосередковано представляють фактори X_5, X_3 і X_4 . Розрахунки кроку процедури аналізу математичної моделі класифікації за даними табл. 2 демонструють зміст ПКР формування моделей класифікації на основі редукції та каппа статистики.

Розглянемо особливість математичної моделі класифікації табл. 1, з урахуванням вимог завдання із визначення авторства україномовних творів (ЗАТ), приведеного у [6, 7]. У нашому варіанті завдання ЗАТ на підставі сукупності ознак, які характеризують набори текстів і утворюють класи або зразки моделі (представляють особистий авторський профіль), необхідно визначити клас, до якого належить новий текст (представлений закодованим вектором ознак). У

завданні система ознак у всіх авторів однакова, а число наборів (творів авторів) може бути різним, не досить значним. При тому певні дані профілю можуть бути «збуреними» або навіть відсутніми. Через те, що окремі твори можуть відноситися до різних періодів, жанрів тощо, формування одного окремого шаблону для кожного із авторів представляє певну проблему. Тому в наведеній моделі класифікації текстів кожному автору відповідає кілька зразків/шаблонів у формі векторів нечітких величин (значень функцій приналежності), що демонструє табл.1. В ній класи K_i – це набори закодованих ознак шаблонів для окремих творів одного автора, серед яких також може бути і шаблон узагальнених показників, «профілю». За результатами досліджень було встановлено [6, 7], що для вибору найбільш ефективних і точних шаблонів моделей класифікації авторів україномовних текстів необхідні додаткові дослідження. В них в першу чергу необхідно визначити раціональні розмірності простору класифікації (число ознак текстів). У числових експериментах з аналізу шаблонів на існування надлишкових характеристик були встановлені можливості скоротити набори ознак $X_1 - X_{14}$, виконати процедуру редукції [1, 3]. Наприклад, для певних вхідних творів були визначені спрощені шаблони із 4-х параметрів, які за результатами класифікації відповідали сукупності параметрів-ознак $X_1 - X_{14}$. Дослідження показали необхідність формування шаблонів моделей завдань ЗАТ з урахуванням вимог методу спрощень [3].

Важливість врахування розмірності моделей класифікації показують оцінки граничної розмірності (3), визначені для умов роботи [1]. Так для умов ($L=70$, $\epsilon=0.2$, $h=0,1$) отримано ($m= 14$, $n_0 = 4,43$)/($m= 7$, $n_0 = 6$)/($m= 6$, $n_0 = 6,53$). Далі при ($L=200$, $\epsilon=0,2$, $h=0,2$) отримано ($m= 14$, $n_0 = 14,6$)/($m= 60$, $n_0 = 9,4$); при $\epsilon=0,1$ граничне значення лише $n_0 = 4,5$, а для забезпечення розмірності $n_0 = 58$ потрібно мати $L = 1200$. Розрахунки показують, що підвищення достовірності результату забезпечується в основному обсягом вибірки даних (числом зразків моделі). Такий висновок визначає форми моделей ЗАТ, в яких безпосередньо використовуються закодовані тексти, без узагальнення даних в шаблонах.

Відзначимо особливості завдання ЗАТ і його реалізації на основі нашої моделі нечіткої класифікації. А саме, в ЗАТ немає вимог щодо числа етапів процедури із визначення автора твору. Разом з тим у [7] та подібних такі завдання передбачалося вирішувати за один розрахунок. Також відсутня необхідність формування єдиної моделі класифікації завдань ЗАТ при любых можливих вхідних творах, а також потреба перетворення моделей шаблонів при введення до моделі нових даних, творів. Зазначені особливості процедур

класифікації враховані у процедурах редукції та каппа Коена, приведених у статті.

На основі зазначених вище моделей класифікації і редукції була розроблена програма для інтелектуальної інформаційної технології (ІІТ) управління призначенням/відбором кандидатів на основі наборів їх певних ознак, рис. 6.

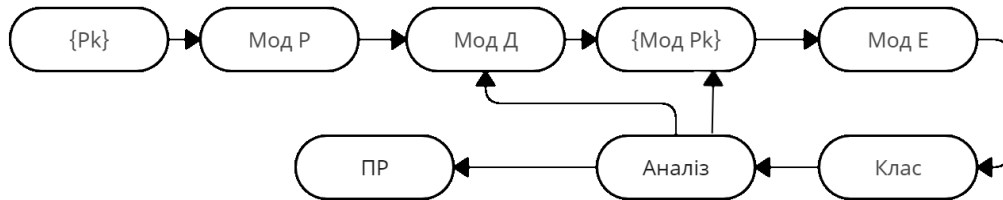


Рисунок 6 - Схема ІІТ управління процесами призначення кандидатів

Блоки структури ІТ визначають наступні процедури: - $\{P_k\}$ – отримання даних портфоліо/резюме кандидатів; - МодР – формування загальної моделі даних портфоліо/резюме конкурсу кандидатів; - МодД – вибір типу моделі представлення даних МодР (нечіткі величини/коефіцієнти впевненості); - $\{Мод P_k\}$ – формування моделей кожного кандидату у форматі МодД; - МодЕ – представлення еталону вимог у форматі МодД; - Клас – реалізація завдання класифікації на моделях $\{Мод P_k\}$, результат P^*_k ; - Аналіз – визначення ступеня однозначності та достовірності результатів класифікації, прийняття результату / визначення необхідності коригування моделей кандидатів $\{Мод P_k\}$ / або коригування загальної моделі даних портфоліо/резюме МодР.

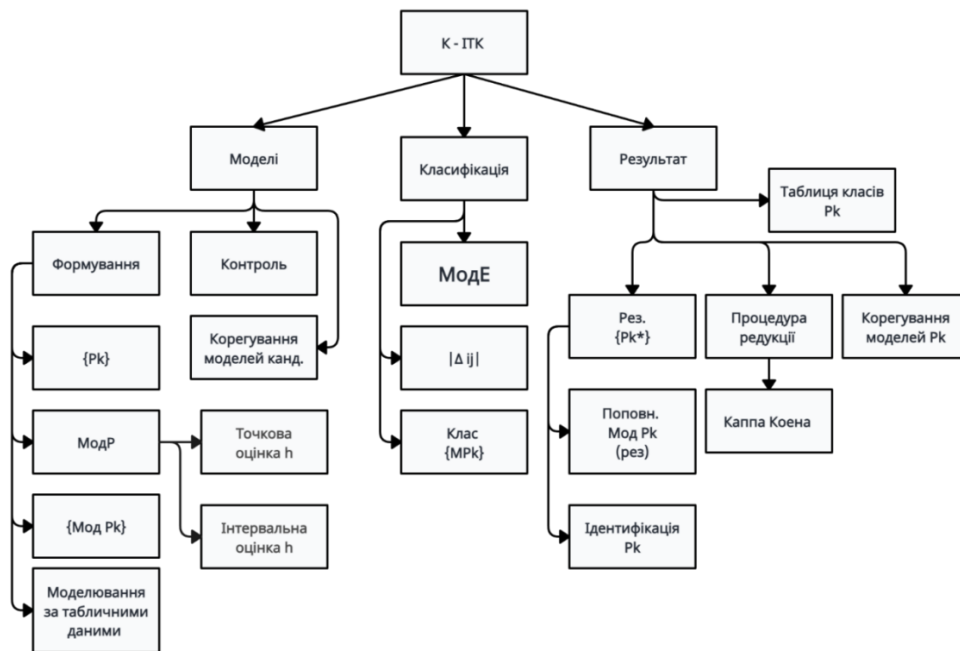


Рисунок 7 - Структура ІІТ програмного комплексу вибору кандидатів

Структурні елементи рис. 7 такі. Б1 Моделі – Формування, контроль, коригування моделей кандидатів {Мод Р_к}; Б2 Класифікація – процедури визначення кандидата-переможця на основі модифікованої мережі Хеммінга; Б3 Результат – представлення результатів вибору кандидата-переможця; Б1.1 Формування - {Р_к}, МодР, {Мод Р_к}; Б1.2 Контроль; Б1.3 Коригування моделей кандидатів; Б2.1 МодЕ – представлення еталону вимог у форматі МодД; Б2.2 Моделі класифікації; Б2.3 Клас – реалізація завдання класифікації на моделях {Мод Р_к}, результат Р*_к; Б3.1 Таблиця класів Р_к; Б3.2 Результат Р*_к; Б3.3 Коригування Мод Р_к; Б1.1.1 {Р_к} – отримання даних портфоліо/резюме кандидатів; Б1.1.2 МодР – формування загальної моделі даних портфоліо/резюме конкурсу кандидатів; Б1.1.3 {Мод Р_к} – формування моделей кожного кандидату у форматі МодД; Б3.2.1 Поповнення Мод Р_к та вивід результатів; Модель і процедура отримання експертних оцінок показників моделей кандидатів {Мод Р_к}. В якості вихідних даних до процедур класифікації на основі МХН надходять закодовані у формі нечітких величин (НВ) або коефіцієнтів впевненості CF(A) вектори ознак моделі класифікації. Такі вектори представляють властивості «шаблонів» {Мод Р_к}, варіантів класифікації, а також векторів вимог – «Е-еталон». Завдання полягає у визначенні «найближчого» до вектору «Е-еталон» шаблону серед всіх {Мод Р_к}. У роботі використано дві моделі та процедури отримання експертних оцінок показників моделей кандидатів {Мод Р_к}. За ними кожному показнику загальної моделі даних портфоліо/резюме МодР призначається (експертом/менеджером) оцінка показника кандидата {Р_к} у формі нечітких величин або коефіцієнтів впевненості CF(A). У першій моделі оцінювання МО1 значення показника вводиться безпосередньо, враховуючи область можливих значень НВ і CF(A).

У другій моделі оцінювання МО2 значення показника визначається на основі окремої нечіткої моделі «Показник» (МПок), призначеної для формування значень на основі опитування (експерта/менеджера).

Розроблений програмний комплекс забезпечує автоматизацію завдань інформаційної технології (ІТ) управління призначенням/відбором кандидатів на основі наборів їх певних нечітких ознак, а також коефіцієнтів CF(A) при заданих ймовірнісних показниках достовірності результатів вибору. Зрозуміло, що комплекс рис. 7 реалізує також завдання ЗАТ, якщо дані портфоліо/резюме представляють закодовані україномовні тексти. На рис. 8 зображено діаграму варіантів використання програмного комплексу рис. 7. На діаграмі представлені наступні функції програми: Ф1 – процедури отримання вихідних

даних; Ф2 – функції збереження даних та результатів досліджень; Ф3 – відображення графіків моделей процесів; Ф4 – формування шаблонів процесів; Ф5 – побудова класів шаблонів для різних типів невизначеності; Ф6 – формування нечітких моделей процедур Хеммінга; Ф7 – налаштування моделей класифікації невизначених даних; Ф8 – аналіз моделей класифікації даних; Ф9 – CF(X) процедури Хеммінга; Ф10 – Оцінювання класів шаблонів; Ф11 – Відображення результатів; Ф12 – процедури порівняльного аналізу.

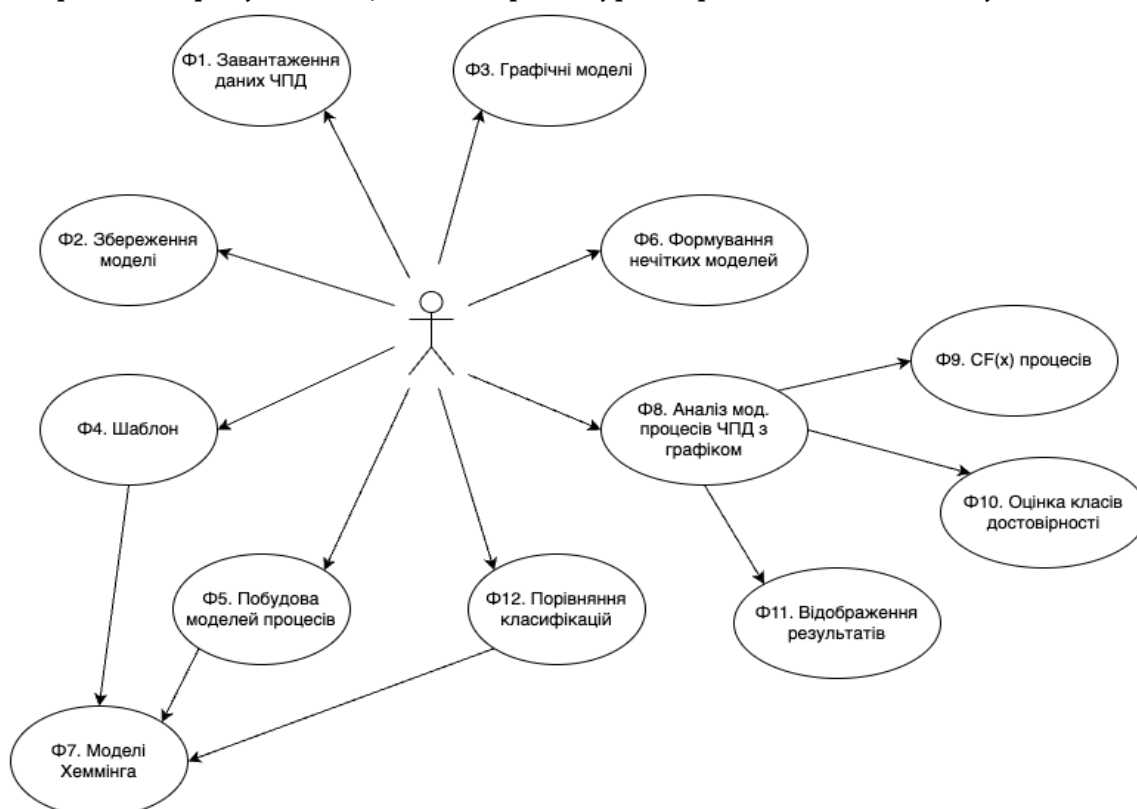


Рисунок 8 - Діаграма варіантів використання програмного комплексу із визначення авторства та призначення/відбір виконавця

Висновки. У статті приведені результати досліджень та розвитку інтелектуальних моделей управління складними системами за умов невизначеності даних на основі процедур, які забезпечують достовірне вирішення завдань з урахуванням оцінки граничної розмірності моделей. Досліджені можливості удосконалення нейронних мереж Хеммінга для класифікації даних у форматах нечітких величин і certainty factor CF(A). Визначені особливості математичної моделі завдань класифікації на основі набору шаблонів ознак. Приведено програмний комплекс інформаційної технології управління призначенням/відбором виконавців, а також визначення авторства україномовних творів на основі класифікації наборів шаблонів із

певних нечітких ознак. Програмний комплекс використовує процедури редукції і каппа статистики.

ЛІТЕРАТУРА

1. Скалозуб В.В., Горячкін В.М., Терлецький І.А., Дудник І.П. Формування моделей класифікації невизначених даних процедурами редукції і каппа статистики. Системні технології. – Дніпро, УДУНТ, 2023. – № 5 (148). – С. 141-155. DOI: 10.34185/1562-9945-5-148-2023-13
<https://journals.nmetau.edu.ua/index.php/st/article/view/1401>
2. Великоіваненко Г. І. Оцінювання рівня економічної безпеки на підґрунті відстані Хеммінга. 2018. URL: <https://core.ac.uk/download/pdf/197269753.pdf>
3. Васільєв В. І. Індукція и редукція в проблемах екстраполяції. Кібернетика і обчислювальна техніка. 1998. Вип. 116. С. 65–81.
4. Колесник А. С., Хайрова Н. Ф. Обґрунтування використання статистики каппа Коена в експериментальних дослідженнях NLP Text Mining. Кібернетика та системний аналіз. Т. 58, № 2. 2022. С. 143–153.
5. Li Min Fu, Shortliffe E. H. The application of certainty factors to neural computing for rule discovery. IEEE Transactions on Neural Networks. 2000. Vol. 11. Iss. 3. P. 647–657. DOI: <https://doi.org/10.1109/72.846736>
6. Скалозуб В. В., Горячкін В. М., Терлецький І. А. Багатопараметричні інтелектуальні процедури діагностування за неповними і збуреними даними // Логістика і транспортна безпека: Проблеми та перспективи розвитку в контексті аналізу сучасних викликів і загроз: матеріали доповідей II Міжнародної науково-практичної конференції. — Дніпро: Середняк Т.К., 2023. С. 42 – 47.
7. Шинкаренко В. І., Демидович І. М. Визначення ознак авторства природно-мовних текстів. Штучний інтелект. 2018. № 3. С. 27–35.
8. Freitag R. M. Ko. Kappa statistic for judgment agreement in Sociolinguistics / Estatística Kappa para concordância de julgamento em Sociolinguística. Revista de Estudos da Linguagem. 2019. Vol. 27, No. 4. P. 1591–1612. DOI: <https://doi.org/10.17851/2237-2083.0.0.1591-1612>
9. Leszek Rutkowski Metody i techniki sztucznej inteligencji. Naukowe PWN, Warszawa, 2005. – 520 p.
10. Haykin S. Neural networks: A Comprehensive Foundation. Prentice hall: New Jersey, 1999. 1103 p.

REFERENCES

1. Skalozub V.V., Horyachkin V.M., Terlets'kyi I.A., Dudnyk I.P. Formuvannya modeley klasyfikatsiyi nevyznachenykh danykh protseduramy reduktsiyi i kappa statys-tyky. Systemni tekhnolohiyi. – Dnipro, UDUNT, 2023. – № 5 (148). – С. 141-155. DOI: 10.34185/1562-9945-5-148-2023-13
2. Velykoivanenko, H. I. (2018). Otsiniuvannya rivnia ekonomichnoi bezpeky na pidgrunti vidstani Khemminha. Retrieved from <https://core.ac.uk/download/pdf/197269753.pdf> (in Ukrainian)
3. Vasilev V. I. Induktsiya i reduktsiya v problemakh ekstrapolyatsii. Cybernetics and Computer Engineering, 1998. Vyp.116, 65-81. (in Russian).
4. Kolesnyk A. S., Khairova N. F. Obgruntuvannya vykorystannia statystyky kappa Koena v eksperymentalnykh doslidzhenniakh NLP Text Mining Cybernetics and system analysis. Vol. 58, No. 2. 2022. P. 143–153.
5. Li Min Fu, Shortliffe E. H. The application of certainty factors to neural computing for rule discovery. IEEE Transactions on Neural Networks. 2000. Vol. 11. Iss. 3. P. 647–657. DOI: <https://doi.org/10.1109/72.846736>
6. Skalozub V.V., Goryachkin V.M., Terletskyi I.A. Bahatoparmetrychni intelektualni protsedury diahnostuvannya za nepovnymy i zburenyymy danymy// Logistics and transport safety: Problems and prospects of development in the context of analysis of modern challenges and threats: materials of reports II International scientific and practical conference. — Dnipro: Serednyak T.K., 2023. P. 42-47.
7. Shinkarenko V. I., Demidovych I. M. Determination of signs of authorship of natural language texts. Artificial Intelligence. 2018. No. 3. P. 27–35.
8. Freitag R. M. Ko. Kappa statistic for judgment agreement in Sociolinguistics / Estatística Kappa para concordância de julgamento em Sociolinguística. Revista de Estudos da Linguagem. 2019. Vol. 27, No. 4. P. 1591–1612. DOI: <https://doi.org/10.17851/2237-2083.0.0.1591-1612>
9. Leszek Rutkowski Metody i techniki sztucznej inteligencji. Naukowe PWN, Warsaw, 2005. – 520 p.
10. Haykin S. Neural networks: A Comprehensive Foundation. Prentice hall: New Jersey, 1999. 1103 p.

Received 12.03.2024.
Accepted 19.03.2024.

Research of intellectual management models based on classification procedures of uncertain data with established requirements of result reliability

For a wide range of complex systems, tasks such as selection of control options for various technological processes, selection of performers for assigned tasks, and determination of authorship are resolved through classification and diagnosis of incomplete data regarding states and conditions of operation. The relevant problems include forming adequate mathematical models of classification procedures and establishing their correctness, completeness, and reliability of results. This article focuses on investigating the properties and development of intellectual management models for complex systems under conditions of data uncertainty based on classification procedures using reduction methods and Cohen's kappa statistics. It is noted that the application of these methods ensures reliable resolution of classification tasks considering the assessment of the maximum model dimensionality. Additionally, the possibilities of improving Hamming neural networks intended for data classification tasks in formats of fuzzy values and certainty factors $CF(A)$ were explored. The features of the proposed enhanced mathematical model for fuzzy classification tasks based on a set of feature templates defining the classes of objects under analysis were identified.

The article also discusses the peculiarities of the mathematical model of classification designed for the task of determining the authorship of Ukrainian-language works (UAW). The characteristics of the UAW task and its implementation based on a fuzzy classification model include the absence of requirements regarding the number of stages in the authorship determination procedure, the unnecessary formation of a unified classification model for UAW tasks for any possible input works, and the absence of the need to transform template models when introducing new data or works into the model. The listed features of classification procedures are accounted for in the reduction and Cohen's kappa procedures outlined in the article.

To implement and study classification tasks of complex system parameters under conditions of uncertain data, appropriate software was developed. The article presents the structure of the software complex for information technology management of performer assignment/selection, as well as the task of determining authorship of Ukrainian-language works based on classification of sets of templates with certain fuzzy features. The software complex utilizes reduction and kappa statistics procedures.

Keywords: template set classification, uncertain data, dimensionality reduction, Cohen's Kappa statistics, fuzzy values, certainty factor $CF(A)$, modified Hamming network, information technology, software, authorship determination, performer assignment/selection.

Скалозуб Владислав Васильович - професор, каф. «Комп'ютерні інформаційні технології», Український державний університет науки і технологій, УДУНТ.

Горячкін Вадим Миколайович – зав. каф. «Комп'ютерні інформаційні технології», Український державний університет науки і технологій, УДУНТ.

Клименко Іван Вікторович – кандидат економічних наук, доцент, доцент кафедри комп'ютерних інформаційних технологій, факультет комп'ютерних систем і технологій, Український державний університет науки і технологій.

Терлецький Ігор Андрійович – аспірант каф. «Комп'ютерні інформаційні технології», Український державний університет науки і технологій, УДУНТ.

Skalozub Vladyslav Vasyliovych– professor, dep. “Computer and Information Technology”, Ukrainian State University of Science and Technology, USUST.

Horiachkin Vadym Mykolayovych – Head of dep. “Computer and Information Technology”, Ukrainian State University of Science and Technology, USUST.

Klymenko Ivan - ass. professor, dep. «Computer and Information Technology», Ukrainian State University of Science and Technology, USUST.

Terlitskyi Ihor Andriyovych– post-graduate student, Dep. “Computer and Information Technology”, Ukrainian State University of Science and Technology, USUST.